Peilen naar de informatiekwaliteit van het net

Het web: een hoog IQ?

Koop je nog maar een soepkip in de supermarkt, dan kan je er van op aan: daar kleeft een stempel van de kwaliteitscontrole op! Ben je echter op zoek naar informatie op het internet, dan kijk je vergeefs uit naar zo'n kwaliteitslabel. Je zal dus zelf moeten beoordelen hoe betrouwbaar de aangeboden informatie is…



(OPZETTELIJK) FOUTIEF?

De New York Times pakte ooit uit met de slogan "On the Internet, nobody knows you're a dog", waarmee men bedoelde dat surfers grotendeels anoniem bleven, en dat je dus nauwelijks kon uitvissen wie je webstek zoal bezocht. Maar eigenlijk geldt ook het omgekeerde. Als surfer is het vaak heel moeilijk om uit te vissen wie nu precies achter de aangeboden informatie schuilgaat - toch een niet onbelangrijke maatstaf om de betrouwbaarheid van de gegevens correct te kunnen inschatten. Daar komt nog bij dat de drempel om iets op het internet te publiceren veel lager is dan - zeg maar – om iets in een gereputeerd tijdschrift afgedrukt te krijgen. Op het net speelt men namelijk z'n eigen uitgever, en komt iemand met zowat alles weg waar hij zin in heeft ('vanity publishing'), zolang het niet indruist tegen de wetgeving. En wat niet met de nationale wetgeving strookt, kan hij wellicht nog wel elders kwijt. Negationistische propaganda (met holocaustontkenning, nvdr) mag dan sedert 1995 in België verboden zijn, maar bijvoorbeeld op Amerikaanse sites bots je nog geregeld op een dergelijke vorm van desinformatie.



Een hond, ook áchter het scherm? (bron: New York Times)



Snel investeren dus maar…

Blijf op je hoede

Zo'n misleiding kan je nog uit politieke beweegredenen verklaren. Maar je moet er ook rekening mee houden dat informatie die via het net wordt uitgestuurd, misleidend kan zijn... gewoon omdat de verspreider dat leuk vindt! Klassiek voorbeeld zijn de hoaxes, e-mailberichtjes die de lezers - onterecht - waarschuwen voor een of ander nieuw supervirus. Soms worden zelfs heuse websites in de lucht gehouden die flagrante onwaarheden bevatten, vaak om er zelf beter van te worden. Zo heeft de Amerikaanse beurscommissie (SEC) enige tijd geleden de namaakwebsite [**www.mcworthle.com**] gelanceerd, precies met de bedoeling gebruikers (lees: investeerders) te wijzen op het gevaar van de talloze sites die opzettelijk foutieve beursinformatie verspreiden - inclusief nagebootste persberichten, aangepaste statistieken, enzovoort.
Maar zelfs als er geen sprake is van opzettelijke misleiding, dan nog moet je de aangeboden informatie kritisch blijven benaderen! Zaak is zo accuraat mogelijk de informatiekwaliteit – ook wel 'IQ' of 'information quality' genoemd – van de gegevens te beoordelen. Verder in het artikel geven we je enkele praktische richtlijnen en vertellen we je welke criteria je kan hanteren om de IQ vast te stellen. Maar voor het zover is... doe je er goed aan je eerst even af te vragen: hoe vind ik eigenlijk die informatie op het net?

ZOEKMACHINES

Diverse onderzoeken laten er geen twijfel over bestaan: *zoekmachines* zitten in de lift, en maar liefst 95 % van alle surfers maakt er geregeld gebruik van. De kans is dus groot dat ook jij via zo'n zoekmachine aan je informatie geraakt. Meteen betekent dit dat diezelfde zoekmachines voor een groot deel bepalen hoeveel én welke gegevens je precies te pakken krijgt! Hoe goed ze daarbij presteren, hangt vooral van twee factoren af. Enerzijds de hoeveelheid informatie die ze weten te ontsluiten (zie ook het kaderstuk: Een web te ver), en anderzijds de manier waarop ze die informatie rangschikken en presenteren. Want zeg nu zelf: hoeveel keer kijk jij verder dan pakweg de 20ste site uit de hitlijst van de zoekmachine? Kampioen wat het aantal geïndexeerde sites betreft, is hoogstwaarschijnlijk zoekrobot Google, die op het moment van schrijven 3,3 miljard webpagina's in z'n databanken heeft verwerkt. Yahoo heeft er heel wat minder, maar dat betekent niet noodzakelijk dat het een slechtere zoekmachine is! De databanken van Yahoo worden namelijk door een team van mensen samengesteld – en niet door een programma of robot zoals bij Google, zodat hier in feite al een informatiefilter zit ingebouwd. Het is natuurlijk nog maar de vraag hoe objectief dit team tewerk gaat bij het (niet) opnemen van sites in de databanken en hoe accuraat de indeling in categorieën is. Het is bijvoorbeeld geen geheim dat heel wat zoekmachines (onder andere AltaVista) nieuw aangemelde sites veel sneller opnemen en up-to-date houden als je voor een of andere betaalformule kiest.

Sponsoring

Meteen zijn we bij de kwaliteit van de samenstelling van de hitlijsten beland. Heel vaak belanden sites van betalende adverteerders namelijk bovenaan de hitlijsten, en soms moet je goed uitkijken om die 'gesponsorde links' van de andere te kunnen onderscheiden. In AltaVista kan je ze met enige moeite herkennen via de onopvallende aanduiding 'Sponsored Matches', terwijl Google ze (voorals-nog?) iets meer afgescheiden presenteert.

AltaVista: sneller en beter voor wie betaalt...

AltaVista: ook hoger voor wie betaalt...

Google laat je ook 'geavanceerd' zoeken.

VAKTAAL

Zoekmachine: Een site die allemaal verwijzingen naar andere sites bevat. Je tikt een trefwoord in en de zoekmachine gaat op het internet zoeken naar sites die verband houden met de door jou ingetypte term. De gevonden sites kan je dan aanklikken om ze te bekijken.

Sponsored Links

Free Christmas Greetings
Create and print Christmas cards
with your photos or our artwork!
www.hp.com/go/homefun
Interest:

Most Unique Gifts
Christmas, collectables, original
special gifts, catalogs available
www.mostuniquegifts.com
Interest:

See your message here...

Eveneens gesponsorde links in Google.

Is sponsoring één criterium waardoor sommige sites een betere rangschikking krijgen in de hitlijsten, dan heb je vaak het raden naar de andere criteria die zoekmachines hanteren om hun hitlijsten te ordenen. Hoe dan ook, nog voor je de informatie op de gevonden websites bekijkt, moet je dus al rekening houden met een subjectieve selectie door de zoekmachines zelf.

Zoektermen

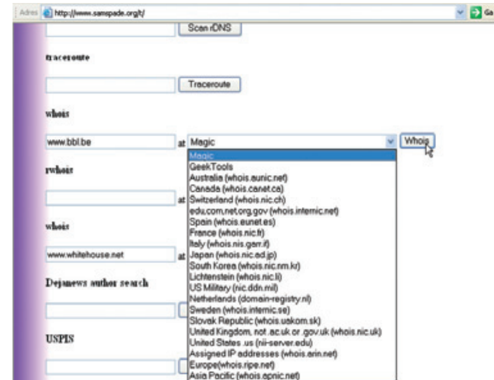
Maar het zijn niet alleen de (algoritmes van de) zoekmachines die instaan voor de kwaliteit van de resulterende hitlijst. Het spreekt voor zich dat wat je als zoekterm(en) intikt mee die kwaliteit bepaalt! Daarbij gaat het niet enkel om het kiezen van de meest geschikte zoektermen, maar ook om de manier waarop je die intikt, en daarvoor is enige kennis van de syntaxregels van de bewuste zoekmachine noodzakelijk. Stel, je wilt informatie sprokkelen rond Willem I. Nu kan je natuurlijk in Google als zoekterm *willem I* intikken, maar dan vergeet je wel dat Google losse cijfers en letters straal negeert. Wil je vermijden dat je talloze onbruikbare Willem-sites moet doorwaden, dan tik je maar beter *"willem I"* of *willem +I* in. De meeste zoekmachines, waaronder Google, voorzien trouwens in een aantal hulpschermen met meer informatie over een optimaal gebruik van zoektermen, én bieden je bovendien een geavanceerd zoekformulier aan waarmee je je zoektocht nog nauwkeuriger kan afbakenen. Ons advies: gebruiken, dat formulier!

INTERNET IQ

Prima, je bent op een website aanbeland – ongeacht of je daar nu via een zoekmachine bent geraakt of niet. Hoe bepaal je nu de IQ van die site? We reiken je meteen een aantal bruikbare criteria aan, maar je moet wel beseffen dat ook het waaróm van je informatiezoektocht een rol speelt. Het maakt wel degelijk verschil uit of je op zoek bent naar kale feiten of bijvoorbeeld (ook) naar opinies. Als je je 'zoekdoel' goed voor ogen houdt, kan je namelijk een stuk sneller inschatten of een bepaalde informatiebron voor jouw doel wel geschikt is.

Geloofwaardigheid

Een basiscriterium is alvast de geloofwaardigheid van de gegevens, en die hangt – zoals je al uit de inleiding kon afleiden – vaak samen met de auteur van de informatie. Anonieme teksten zouden je tot extra voorzichtigheid moeten aanzetten, zeker als ook nog blijkt dat de taal en de stijl waarin de informatie is vastgelegd, te wensen overlaat. Wordt de auteursnaam vermeld, kijk dan niet op de moeite om – bijvoorbeeld via zoekmachines – naar meer informatie over die persoon op zoek te gaan. Ook de *url* van de webpagina kan een stevige hulp zijn om de geloofwaardigheid van de informatie correcter in te schatten. Je mag wellicht meer vertrouwen stellen



Speuren achter de schermen.

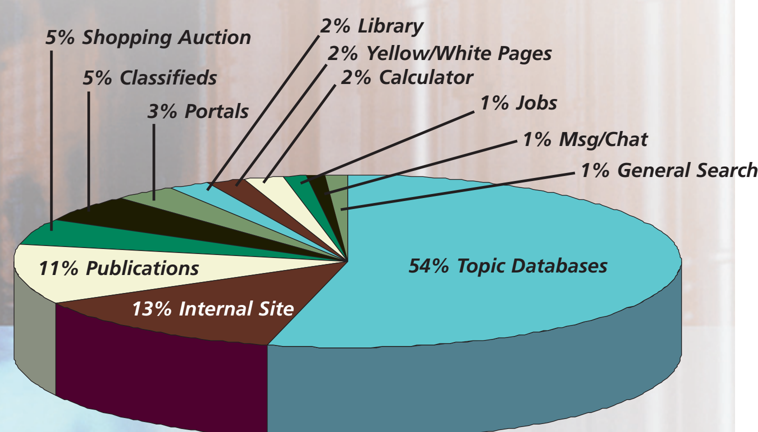
in gegevens die je gevonden hebt op [**www.wwf.org**] dan op [**users.provider.net/a102057/~username**] – tenminste als je informatie zocht naar het wildbestand in het zuidelijk halfrond. Ben je ergens midden in een site beland, en is het je niet meteen duidelijk met welke auteur of organisatie je te maken hebt, dan kan je even een blik op de homepagina van die site werpen. Is daarvoor niet onmiddellijk een geschikte link voorzien, dan kan je de url stelselmatig strippen - van rechts naar links - tot wanneer je aan de eerste slash (/) beland bent. Die afgeslankte url brengt je gewoonlijk wél tot bij de openingspagina, van waaruit je beter kan bepalen om welke organisatie het precies gaat. Hoewel, ook hier is het soms opletten geblazen: 'url-grabbers' (ook wel 'page-jackers' genoemd) durven zich onrechtmatig de url van een bestaande organisatie toe te eigen, óf ze gebruiken gewoon een url die verwarring schept. Zeg nu zelf: [**www.whitehouse.gov**], [**www.whitehouse.com**] en [**www.whitehouse.net**...] wat is nu de "echte" site? Geraak je er niet meteen wijs uit, dan kan je nog altijd een beroep doen op een zogenaamde Whois-query. Die vertelt je dan wie er precies achter de geregistreerde *domeinnaam* schuilgaat. Op [**www.samspade.org/t**] bijvoorbeeld kan je gratis zo'n wie is-speurtocht uitvoeren.



Wil het echte Witte Huis nu opstaan?

EEN WEB TE VER

Het internet is een gigantische informatievergaarbak, zozeer zelfs dat we met 'information overload' of 'infoglut' af te rekenen krijgen. We worden gewoonweg overstelpt met massa's gegevens en kunnen die nog nauwelijks verwerken.
En toch… zou Google met z'n 3,3 miljard pagina's naar schatting slechts 25 % van de vrij toegankelijke webpagina's verwerkt hebben! En volgens sommige bronnen (zie [**www.brightplanet.com/technology/deepweb.asp**]) zouden meer verborgen webpagina's maar liefst 500 keer méér informatie bevatten! In dat geval zou Google niet meer dan 1 op 2.000 webdocumenten ontsloten hebben. Komt daar nog bij dat het web dagelijks met enkele miljoenen documenten aangroeit, zodat de kloof alleen maar groter dreigt te worden. En zelfs als je de resultaten van alle bekende zoekmachines bij elkaar legt, kom je niet veel hoger uit dan 35 % ontsluiting, aangezien de databanken van die zoekmachines elkaar grotendeels overlappen.
Al die extra informatie huist in het zogenaamde diepe web. Dat zijn in de eerste plaats pagina's die dynamisch gegenereerd worden: ze worden op het moment zelf samengesteld uit gegevens die zich in allerlei databanken bevinden, bijvoorbeeld op grond van de zoekopdracht die een gebruiker heeft ingetikt. Los daarvan weten heel wat zoekmachines zich op dit moment evenmin raad met webdocumenten als doc, ppt, xls en pdf, en die worden alsmaar talrijker. Google kan bijvoorbeeld wel overweg met pdf-bestanden, maar enkel het eerste stukje van zo'n document wordt door de zoekrobot effectief ingelezen en geïndexeerd. Deze gedeeltelijke indexering geldt overigens ook voor gewone webpagina's: de meeste zoekrobots bezoeken namelijk zelden webpagina's die meer dan drie linkniveaus diep huizen. Ze bezoeken bijvoorbeeld wel de homepagina, én pagina's waar die naar linkt, maar nog een niveau verder houden ze het meestal voor bekeken. Ten slotte sluiten heel wat webmasters bewust een stuk van hun site af voor zoekrobots (bijvoorbeeld via een *metatag* als: `<meta name="robots" content="noindex,nofollow">`, of schermen ze onderdelen af via een gebruikerswachtwoord.
Intussen worden wel allerlei pogingen ondernomen om ook dat diepe web wat verder te ontsluiten. Je kan alvast even proberen of je méér kan opvissen via de zoekdiensten van CompletePlanet [**www.completeplanet.com**], Invisible Web [**www.invisible-web.net**] en Profusion [**www.profusion.com**] – maar ons wisten ze alvast niet te overtuigen! Het ziet er dus niet naar uit dat het diepe web zijn geheimen morgen al zal prijsgeven…



Het diepe web, geordend volgens inhoud. (bron: BrightPlanet)

Accuraatheid

De accuraatheid van de informatie is een tweede criterium. Een belangrijke parameter kan de datum zijn waarop die gegevens zijn opgenomen, vooral als die informatie van vluchtige aard is. Het is bijvoorbeeld niet omdat een bekend persoon 10 jaar geleden een uitspraak deed, dat die nog steeds geldig is! De accuraatheid van een informatiebron hangt ook vaak samen met het doelpubliek: aan wetenschappelijke publicaties kan je gewoonlijk andere eisen stellen dan aan informatie die in eerste instantie op een erg jeugdig publiek mikt. Tracht je tevens een goed oordeel te vormen over de diepgang (hoeveel details bevat de informatie) en reikwijdte (hoeveel deelaspecten komen aan bod) van de gegevens. Een goed gestructureerde site zou je overigens voldoende navigatiemogelijkheden moeten bieden om daar snel je weg in terug te vinden.
Hoed je ook voor 'verborgen boodschappen' (zoals je die vaak op dotcoms vindt): wat een leerrijke uiteenzetting lijkt over het gevaar van virussen, is wellicht bedoeld als een poging om je een antivirusprogramma van de producent aan te smeren. Dat doet niet noodzakelijk iets af aan de kwaliteit van de informatie, maar het hoort je wel extra kritisch te stemmen.

Redelijkheid

Meteen komen we bij een derde criterium: de redelijkheid. Hiertoe behoort ook de (graad van) objectiviteit van de aangeboden informatie. Zuivere objectiviteit is weliswaar nauwelijks haalbaar, maar komt de inhoud je als tendentieus over, dan moet je extra uit je doppen kijken. Dat geldt des te meer als het om erg persoonlijke inbreng gaat (zoals in weblogs) en *nieuwsgroep*postings (zoals op [**http://groups.google.be**]). Hoop je via dergelijke bronnen ook zuiver feitenmateriaal op te snorren, dan doe je er goed aan op zoek te gaan naar andere bronnen (op het internet) die de informatie van je eerste bron kunnen bevestigen. Zo'n vergelijkende controle bouw je overigens best altijd in, ook als je meer objectieve sites consulteert! Het is natuurlijk handig als de geraadpleegde site zelf een (aanklikbare) lijst naar andere, externe pagina's aanbiedt die dezelfde topics behandelen. Maar ook hier is het opletten: vergewis je ervan dat die externe pagina's toch niet op een of andere manier aan de oorspronkelijke pagina gelieerd zijn.

— Toon Van Daele —

VAKTAAL

Domeinnaam: Een unieke naam die verwijst naar een internetadres. De naam bestaat minimaal uit twee delen gescheiden door een punt, zoals: clickx-magazine.be. Alle domeinnamen zijn gekoppeld aan een zogenaamd IP-nummer, maar zo'n ingewikkelde nummercombinatie laat zich moeilijk onthouden.

Metatag: Stukje html-code op een webpagina waarin je allerlei informatie over je html-document kwijt kan. Sommige zoekrobots kunnen die informatie lezen.

Nieuwsgroepen: Nieuwsgroepen kan je vergelijken met (virtuele) prikborden die berichten verzamelen rond bepaalde thema's. Op het internet vind je letterlijk tienduizenden nieuwsgroepen waar je berichten kan lezen en zelf berichten naar kan versturen.

Url: Staat voor Uniform Resource Locator. Een uitdrukking die het adres van een website op het internet aanduidt.